Systematic alanine scanning of PAX8 paired domain reveals functional importance of the N-subdomain

Short title: Alanine scanning mutagenesis of PAX8

Authors:

Megumi Iwahashi [1,2] and Satoshi Narumi [1]

Affiliations:

[1] Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan

[2] Department of Pediatrics, The Jikei University School of Medicine, Tokyo 105-8461, Japan.

Correspondence:

Satoshi Narumi, M.D., Ph.D.

Department of Molecular Endocrinology, National Research Institute for Child Health and Development

2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan.

TEL: +81-3-3416-0181

E-mail: narumi-s@ncchd.go.jp

Key words: paired domain, PAX, PAX8, alanine scanning mutagenesis

Funding: This work was supported by JSPS KAKENHI (Grant Number 15K09630) from the Japan Society for the Promotion of Science.

Word count: abstract, 249; text, 2,535; figures 2; table 2.

Abstract

Thyroid-specific transcription factor PAX8 has an indispensable role in the thyroid gland development, which is evidenced by the facts that *PAX8/Pax8* mutations cause congenital hypothyroidism in humans and mice. More than 90% of known *PAX8* mutations were located in the paired domain, suggesting the central role of the domain in exerting the molecular function. Structure-function relationships of PAX8, as well as other PAX family transcription factors, have never been investigated in a systematic manner. Here, we conducted the first alanine-scanning mutagenesis study, in which 132 alanine variants located in the paired domain of PAX8 were created and systematically evaluated *in vitro*. We found that 76 alanine variants (55%) were loss of function (LOF) variants (defined by <30% activity as compared with wildtype PAX8). Importantly, the distribution of LOF variants were skewed, with more frequently observed in the N-subdomain (65% of the alanine variants in the N-subdomain) than in the C-subdomain (45%). Twelve out of 13 alanine variants in residues that have been affected in patients with congenital hypothyroidism were actually LOF, suggesting that the alanine scanning data can be used to evaluate the functional importance of mutated residues. Using our *in vitro* data, we tested the accuracy of seven computational algorithms for pathogenicity prediction, showing that they are sensitive but not specific to evaluate on the paired domain alanine variants. Collectively, our experiment-based data would help better understanding of the structure-function relationships of the paired domain, and would provide a unique resource for pathogenicity prediction of future *PAX8* variants.

Introduction

Organ-specific transcription factors play pivotal roles in organogenesis, organ growth, and maintenance of organ functions. As for the thyroid gland, three thyroid-specific transcription factors (TTF), namely PAX8 (encoded by *PAX8*) (Plachov *et al.* 1990, Zannini *et al.* 1992), TTF-1 (encoded by *NKX2-1*) (Guazzi *et al.* 1990, Lazzaro *et al.* 1991), TTF-2 (encoded by *FOXE1*) (Zannini *et al.* 1997), and HEX (encoded by *HHEX*) (Crompton *et al.* 1992, Thomas *et al.* 1998), have been known to play such roles.

PAX8 is a member of the PAX gene family, which is characterized by the presence of DNA-binding paired domain. In mice and human, *Pax8/PAX8* is expressed in the thyroid from prenatal period to adulthood (Plachov *et al.* 1990). In cultured cell lines, PAX8 directly regulates the transcription of thyroid-specific genes, such as thyroglobulin (Tg), thyroid peroxidase (TPO) (Di Palma *et al.* 2003) and the sodium iodide symporter (Ohno *et al.* 1999). Genetically-engineered Pax8-deficient mice show severe thyroid hypoplasia due to defective proliferation and survival of thyroid precursor cells (Mansouri *et al.* 1998), indicating the indispensable role of Pax8 in the early organogenesis. In humans, heterozygous *PAX8* mutations cause congenital hypothyroidism (CH) with autosomal dominant inheritance. To date, 33 mutation-carrying families harboring a total of 23 distinct *PAX8* mutations have been described (Al Taji *et al.* 2007, Carvalho *et al.* 2013, Congdon *et al.* 2001, de Sanctis *et al.* 2004, Di Palma *et al.* 2010, Grasberger *et al.* 2005, Hermanns *et al.* 2013, Jo *et al.* 2010, Komatsu *et al.* 2001, Liu *et al.* 2017, Lof *et al.* 2016, Macchia *et al.* 1998, Meeus *et al.* 2004, Narumi *et al.* 2012, Narumi *et al.* 2010, Narumi *et al.* 2011, Ramos *et al.* 2014, Srichomkwun *et al.* 2016, Tonacchera *et al.* 2007, Vilain *et al.* 2001, Zou *et al.* 2015). Clinical phenotypes of the mutation carriers are variable, ranging from overt CH with severe thyroid hypoplasia to subclinical hypothyroidism

with a normal-sized thyroid (Supplementary Table 1).

Previously reported experimentally-verified PAX8 mutations have been found exclusively in the paired domain, except for one truncating mutation (p.Thr 277*) located outside the domain (de Sanctis *et al.* 2004). This fact implies the importance of the paired domain in exerting the molecular function of PAX8. Nonetheless, genetic variants located in functionally important domains do not necessarily result in loss of function (LOF) sufficient for disease onset. Hence, when a novel variant is found in a patient, it is necessary to presume its pathogenicity. To this end, a handful of computational algorithms, such as PolyPhen-2 (Adzhubei *et al.* 2010) and SIFT (Kumar *et al.* 2009), have been used. However, the accuracy of these *in silico* algorithms remains around about 70% (Thusberg *et al.* 2011), which is not enough reliable to be used in clinical genetic diagnosis. In this present study, systematic alanine scanning mutagenesis was performed to determine which residues are functionally important. Alanine substitution eliminates side chain interactions without altering main-chain conformation, enabling to assess the contribution of specific residue on the function. We performed systematic alanine scanning mutagenesis that targeted all 132 non-alanine residues of the paired domain of PAX8, and assessed the effects of amino acid alterations *in vitro*. We also compared *in silico* algorithms accuracy with use of our *in vitro* dataset.

Materials and Methods

Plasmids

We used simian virus 40 promoter-driven effector plasmids carrying human PAX8 cDNA that has been

previously described (Narumi *et al.* 2010). We created a total of 132 alanine-substituted, the smallest chiral amino acid, variant PAX8 [Met1Ala to Gln137Ala; five alanine residues (Ala^{19}, Ala^{38}, Ala^{84}, $Ala^{104}$and Ala^{113}) were excluded] using PrimeSTAR Mutagenesis Basal Kit (TaKaRa Bio Inc, Shiga, Japan). We confirmed each alanine mutation substitution by direct sequencing. Alanine scanning enables quick determination of each individual amino acid's contribution to the protein function. The transcriptional activities of PAX8 proteins [wildtype (WT) or alanine variants] were assessed with firefly luciferase reporters that contain the promoter sequence of the human Tg gene corresponding to -284/+39 region (*TG*-luc) (Narumi *et al.* 2010), or the promoter sequence of the rat TPO gene corresponding to -1/+426 region (*Tpo*-luc) (Di Palma *et al.* 2003).

Cell culture and transfection

HeLa cells were maintained in DMEM supplemented with 100 U/mL penicillin, 100 μg/mL streptomycin, and 10% fetal bovine serum. Transient transfection was performed with the Lipofectamine 3000 reagent (Thermo Fisher Scientific, Waltham, MA, USA). Cells grown in 96-well plates with 70-80% confluence were transfected with 90 ng of each luciferase reporter (*TG*-luc or *Tpo*-luc), and 10 ng of each effector plasmid (empty vector, WT-PAX8 or alanine variant PAX8). Forty-eight hours after transfection, we measured luciferase activities using ONE-Glo Luciferase Assay System (Promega, Madison, WI) according to the manufacture’s instruction. Luciferase activities were represented relative to the activity obtained by transfection of WT-PAX8 (set to 100%) and empty vector (set to 0%). Experiments were conducted in

quadruplicate, and were repeated at least three times. The activity data were expressed as mean±SEM. Based on the transactivating capacities, the following terms were defined: profound LOF, less than 10.0% activity; moderate LOF, 10.0–29.9% activity; minimal LOF, 30.0-69.9% activity; functionally neutral, 70.0-119.9% activity; hyperfunctioning, equal or more than 120% activity. We considered variants with profound LOF or moderate LOF as disease-causing variants (described as significant LOF).

Three-dimensional modeling

The three-dimensional modeling structure of PAX8–DNA complex has not been determined to date. Instead, we used crystal structure data of the PAX5-DNA complex to display the positions of the variants, considering the high protein sequence similarity (94% identical in the paired domain) between PAX8 and PAX5. The crystal structure data of PAX5–DNA complex was obtained from protein data bank (ID 1MDM; http://www.rcsb.org/pdb), and was used as a template to visualize the effects of alanine substitutions. The pictures were produced with PyMOL (http://www.pymol.org). Residues that were not identical between PAX8 and PAX5 were not shown. The effect of alanine substitution in the pairing region of PAX8 was shown as color-coded spheres: red, profound LOF; orange, moderate LOF; yellow, minimal LOF; gray, functionally neutral; and blue, hyperfunctioning. Presumed effects of alanine substitution were classified into five categories: (i) loss of hydrogen bond(s) to target DNA; (ii) loss of van der Waals contact(s) to target DNA; (iii) loss of hydrogen bond(s) within PAX8, (iv) loss of van der Waals contact(s) within PAX8; and (v) no recognizable effect on contact to target DNA or within PAX8 (“free sidechain” residues). We used

PyMOL to predict hydrogen bonds with a default setting van der Waals contacts were defined based on the distance (3.50 to 3.99 Å) between the Ala sidechains to other carbon molecules (target DNA or other residues of PAX8).

Computational prediction of pathogenicity of PAX8 alanine variants

The pathogenicity of 132 alanine variants were assessed by following seven computational algorithms: FATHMM (Shihab *et al.* 2013), MutationAssessor (Reva *et al.* 2011), MutationTaster (Schwarz *et al.* 2010), PolyPhen-2 (Adzhubei *et al.* 2010), PROVEAN (Choi *et al.* 2012), SIFT (Kumar *et al.* 2009), and VEST-4 (Carter *et al.* 2013). The accuracy of computational prediction tools was assessed by ROC curves. In this study, a gold standard was defined by the results of *in vitro* functional assays, in which variants with significant LOF (*i.e.,* transactivating capacity less than 30%) were considered to be deleterious.

Results

Transactivating capacities of alanine variants

PAX8 is a 450 amino-acids protein that contains a paired domain of Gly^{9} to Gln^{137}. To systematically assess the effect of amino acid substitutions in the paired domain and its N-terminal region (Met^{1} to Ser^{8}), 132 non-alanine residues were substituted to alanine one by one, and transcriptional activities of each mutant were assessed with two luciferase reporters (*TG*-luc and *Tpo*-luc) (Supplementary Table 2). There was a significant correlation between the transactivating capacities for *TG*-luc and *Tpo*-luc among the 132 alanine

variants ($R^2 = 0.66$, $P<0.001$; Supplementary Fig. 1). Discordance in transactivating capacities between the two reporters was observed in only one alanine variant Val75Ala (*TG*-luc 65.8±12.3%, *Tpo*-luc 127.0±17.2%). For simplicity, descriptions of transactivating capacities below are based on ones measured with *TG*-luc.

Profound LOF was observed in 46 alanine variants (35% of total alanine variants), while moderate LOF was observed in 30 alanine variants (23%) (Fig. 1A and 1B). Eight alanine variants (7%) were considered to be functionally neutral. One alanine variant (Glu67Ala) showed slightly elevated transactivating capacities (139±7.9% activity relative to WT-PAX8).

Structure-function relationships of the paired domain of PAX8

The paired domain consists of two β-sheets (β1 and β2) and six α-helices (α1 to α6) (Fig. 1A). Based on the crystal structure data, paired domains can be subdivided to three segments: N-terminal subdomain (β1, β2, α1 to α3), C-terminal subdomain (α4 to α6), and the linker polypeptide between the two subdomains (Fig. 1B). As for 66 alanine variants in the N-subdomain, 30 (45%) were profound LOF and 13 (20%) were moderate LOF (Fig. 1C). As for 50 alanine variants in the C-subdomain, nine (18%) were profound LOF and 14 (28%) were moderate LOF (Fig. 1D). The effects of alanine substitution in the linker polypeptide were generally modest, except for Gly77, Gly78 and Ser79 that contact with DNA at the bottom of minor groove (Fig. 1E).

Based on the crystal structure data, we classified the effects of the alanine substitution into five

categories according to the presence or absence of contact(s) to the target DNA or within PAX8 (Table 1). Contacts to the target DNA, which were chiefly via hydrogen bonds, were found in twelve and five residues in N- and C-subdomains, respectively. About 70% of alanine substitutions of these residues caused significant LOF. Intramolecular contacts (hydrogen bonds or van der Waals contacts) were seen in 25 and 29 residues in N- and C-subdomains, respectively. About 60% of alanine substitutions of the residues caused significant LOF in both of the two domains. “Free sidechain” residues were observed in 16 and 11 residues in N- and C-subdomains, respectively. Eleven out of 16 “free sidechain” residues in N-subdomain caused, when mutated to alanine, significant LOF, while only three out of 11 did in C-subdomain (P=0.036 by Fisher exact test).

Alanine substitution on the residues affected in CH patients

To date, a total of 16 distinct CH-causing missense *PAX8* mutations affecting 13 residues have been described in the paired domain (Al Taji *et al.* 2007, Carvalho *et al.* 2013, Congdon *et al.* 2001, Di Palma *et al.* 2010, Grasberger *et al.* 2005, Hermanns *et al.* 2013, Lof *et al.* 2016, Macchia *et al.* 1998, Meeus *et al.* 2004, Narumi *et al.* 2012, Narumi *et al.* 2010, Ramos *et al.* 2014, Srichomkwun *et al.* 2016, Vilain *et al.* 2001). Out of the 13 residues, 12 caused significant LOF when substituted to alanine (Table 2). The only exception was Gln40Ala (corresponding human mutation, Gln40Pro), which showed minimal LOF (61.0±1.8% activity).

Comparison of computational algorithms

With assuming *in vitro* transactivating capacities as a gold standard test, we compared accuracy of seven computational algorithms (FATHMM, MutationAssessor, MutationTaster, PolyPhen-2, PROVEAN, SIFT, and VEST-4) in predicting the pathogenicity of 132 PAX8 alanine variants (Supplementary Table 2). The area under the ROC curve values of the seven algorithms were 0.65±0.09 (mean±SD) (Fig. 2). Highest area under the ROC curve was scored in MutationAssessor with 0.76. When default cut-off threshold was applied, all seven computational algorithms showed high sensitivity (0.95±0.04) and very low specificity (0.14±0.15). When optimal cut-off values based on the ROC curves were used, the sensitivity and specificity were 0.62±0.21 and 0.63±0.14, respectively (Supplementary Fig. 2A-G).

Discussion

In this study, we conducted the first systematic alanine-scanning mutagenesis targeting the paired domain of a PAX transcription factor. Our findings would provide unique insights into structure-function relationships of the paired domain of PAX8, and possibly other PAX gene family transcription factors, since amino-acid sequences (and three-dimensional structures probably) are well conserved (Supplementary Fig. 3).

Three segments (N-subdomain, linker polypeptide and C-subdomain) were recognizable in the paired domain of PAX8. Among 76 alanine variants with significant LOF, 43, 10 and 23 were located in the N-subdomain, linker polypeptide and C-subdomain, respectively. Relatively severe effects of alanine substitutions on the N-subdomain indicate the functional importance of the subdomain. The assumption

agrees well with following three facts: (i) the N-subdomain is better conserved among PAX family genes than the C-subdomain (Supplementary Fig. 3); (ii) vast majority of CH-causing *PAX8* mutations have been observed in the N-subdomain (Fig. 1A, Supplementary Table 3); (iii) similarly, mutations have been preferentially observed in the N-subdomain in PAX3 defect (Waardenburg syndrome), PAX6 defect (Aniridia) and PAX9 defect (Oligodontia or hypodontia), while mutations show broader distribution in PAX2 defect (Renal-coloboma syndrome) (Supplementary Table 3) (Schimmenti 2011). We could recognize two structural/functional differences between N- and C-subdomains. One is the numbers of contacts between PAX8 and the target DNA (12 contacts in N-subdomain; 5 contacts in C-subdomain). N-subdomain is expected to bind to the target DNA more tightly. The other difference is related to "free sidechain" residues. Among 16 “free sidechain” residues in N-subdomain, 69% caused, when mutated to alanine, significant LOF, while only 27% in C-subdomain. Relatively severe effects of alanine substitutions on “free sidechain” residues in N-subdomain imply the yet unknown function(s) of these residues. One possible explanation is interaction(s) between the PAX8 protein and other transcription factor(s). PAX5, a closely related transcription factor of PAX8, is known to bind to Ets by its N- subdomain to form a complex. PAX8 might have such partner molecule(s), although none have been identified so far.

Alanine substitutions of 13 residues that have been affected in human CH patients caused significant LOF, except for Gln^{40}. This failure in prediction is probably explained by the nature of substitutions: Gln to Ala change affects the side chain only, while Gln to Pro change affects both the main chain and side chain. We suppose that alanine scanning-based data will be a useful resource for pathogenicity prediction of future novel PAX8 variants, although the nature of amino acid substitution should also be considered.

In this study, we showed that widely used computational algorithms have two clear deficits in pathogenicity prediction of PAX8 variants. First, area under the ROC curve values of the algorithms were generally low, indicating insufficient performance of them. Among the seven tested algorithms, MutationAssessor worked best. Second, default cut-off thresholds for prediction were set inappropriately low in all seven algorithms. This would result in false positive results. We stress that caution for "overdiagnosis" is needed when clinicians interpret the results of computational algorithms with default cut-off thresholds.

In summary, we report alanine-scanning mutagenesis that targeted all 132 non-alanine residues of the paired domain of PAX8. Our works not only contribute to better understanding of the structure-function relationships of the paired domain, but also provides a unique resource for evaluating the pathogenicity of novel *PAX8* variants identified in future CH patients.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This work was supported by JSPS KAKENHI (Grant Number 15K09630) from the Japan Society for the Promotion of Science.

Acknowledgements

The authors would like to thank Mariastella Zannini for providing us the *Tpo*-luc construct. We thank Prof. Hiroyuki Ida of Department of Pediatrics, Jikei University School of Medicine for fruitful discussion. This work was supported by JSPS KAKENHI (Grant Number 15K09630) from the Japan Society for the Promotion of Science.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS & Sunyaev SR 2010 A method and server for predicting damaging missense mutations. *Nat Methods* **7** 248-249. (http://doi.org/10.1038/nmeth0410-248)

Al Taji E, Biebermann H, Limanova Z, Hnikova O, Zikmund J, Dame C, Gruters A, Lebl J & Krude H 2007 Screening for mutations in transcription factors in a Czech cohort of 170 patients with congenital and early-onset hypothyroidism: identification of a novel PAX8 mutation in dominantly inherited early-onset non-autoimmune hypothyroidism. *Eur J Endocrinol* **156** 521-529. (http://doi.org/10.1530/EJE-06-0709)

Carter H, Douville C, Stenson PD, Cooper DN & Karchin R 2013 Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **14 Suppl 3** S3. (http://doi.org/10.1186/1471-2164-14-S3-S3)

Carvalho A, Hermanns P, Rodrigues AL, Sousa I, Anselmo J, Bikker H, Cabral R, Pereira-Duarte C, Mota-Vieira L & Pohlenz J 2013 A new PAX8 mutation causing congenital hypothyroidism in three generations of a family is associated with abnormalities in the urogenital tract. *Thyroid* **23** 1074-1078. (http://doi.org/10.1089/thy.2012.0649)

Choi Y, Sims GE, Murphy S, Miller JR & Chan AP 2012 Predicting the functional effect of amino acid substitutions and

230 indels. *PLoS One* **7** e46688. (http://doi.org/10.1371/journal.pone.0046688)

231 Congdon T, Nguyen LQ, Nogueira CR, Habiby RL, Medeiros-Neto G & Kopp P 2001 A novel mutation (Q40P) in

232 PAX8 associated with congenital hypothyroidism and thyroid hypoplasia: evidence for phenotypic variability in

233 mother and child. *J Clin Endocrinol Metab* **86** 3962-3967. (http://doi.org/10.1210/jcem.86.8.7765)

234 Crompton MR, Bartlett TJ, MacGregor AD, Manfioletti G, Buratti E, Giancotti V & Goodwin GH 1992 Identification

235 of a novel vertebrate homeobox gene expressed in haematopoietic cells. *Nucleic Acids Res* **20** 5661-5667.

236 de Sanctis L, Corrias A, Romagnolo D, Di Palma T, Biava A, Borgarello G, Gianino P, Silvestro L, Zannini M &

237 Dianzani I 2004 Familial PAX8 small deletion (c.989_992delACCC) associated with extreme phenotype variability.

238 *J Clin Endocrinol Metab* **89** 5669-5674. (http://doi.org/10.1210/jc.2004-0398)

239 Di Palma T, Nitsch R, Mascia A, Nitsch L, Di Lauro R & Zannini M 2003 The paired domain-containing factor Pax8

240 and the homeodomain-containing factor TTF-1 directly interact and synergistically activate transcription. *J Biol*

241 *Chem* **278** 3395-3402. (http://doi.org/10.1074/jbc.M205977200)

242 Di Palma T, Zampella E, Filippone MG, Macchia PE, Ris-Stalpers C, de Vroede M & Zannini M 2010 Characterization

243 of a novel loss-of-function mutation of PAX8 associated with congenital hypothyroidism. *Clin Endocrinol (Oxf)* **73**

244 808-814. (http://doi.org/10.1111/j.1365-2265.2010.03851.x)

245 Grasberger H, Ringkananont U, Lefrancois P, Abramowicz M, Vassart G & Refetoff S 2005 Thyroid transcription factor

246 1 rescues PAX8/p300 synergism impaired by a natural PAX8 paired domain mutation with dominant negative

247 activity. *Mol Endocrinol* **19** 1779-1791. (http://doi.org/10.1210/me.2004-0426)

248 Guazzi S, Price M, De Felice M, Damante G, Mattei MG & Di Lauro R 1990 Thyroid nuclear factor 1 (TTF-1) contains

249 a homeodomain and displays a novel DNA binding specificity. *EMBO J* **9** 3631-3639.

250 Hermanns P, Grasberger H, Cohen R, Freiberg C, Dorr HG, Refetoff S & Pohlenz J 2013 Two cases of thyroid
251 dysgenesis caused by different novel PAX8 mutations in the DNA-binding region: in vitro studies reveal different
252 pathogenic mechanisms. *Thyroid* **23** 791-796. (http://doi.org/10.1089/thy.2012.0141)
253 Jo W, Ishizu K, Fujieda K & Tajima T 2010 Congenital Hypothyroidism Caused by a PAX8 Gene Mutation Manifested
254 as Sodium/Iodide Symporter Gene Defect. *J Thyroid Res* **2010** 619013. (http://doi.org/10.4061/2010/619013)
255 Komatsu M, Takahashi T, Takahashi I, Nakamura M, Takahashi I & Takada G 2001 Thyroid dysgenesis caused by
256 PAX8 mutation: the hypermutability with CpG dinucleotides at codon 31. *J Pediatr* **139** 597-599.
257 (http://doi.org/10.1067/mpd.2001.117071)
258 Kumar P, Henikoff S & Ng PC 2009 Predicting the effects of coding non-synonymous variants on protein function
259 using the SIFT algorithm. *Nat Protoc* **4** 1073-1081. (http://doi.org/10.1038/nprot.2009.86)
260 Lazzaro D, Price M, de Felice M & Di Lauro R 1991 The transcription factor TTF-1 is expressed at the onset of thyroid
261 and lung morphogenesis and in restricted regions of the foetal brain. *Development* **113** 1093-1104.
262 Liu S, Wang X, Zou H, Ge Y, Wang F, Wang Y, Yan S, Xia H & Xing M 2017 Identification and characterization of
263 novel PAX8 mutations in Congenital Hypothyroidism(CH) in a Chinese population. *Oncotarget* **8** 8707-8716.
264 (http://doi.org/10.18632/oncotarget.14419)
265 Lof C, Patyra K, Kuulasmaa T, Vangipurapu J, Undeutsch H, Jaeschke H, Pajunen T, Kero A, Krude H, Biebermann H,
266 *et al.* 2016 Detection of Novel Gene Variants Associated with Congenital Hypothyroidism in a Finnish Patient
267 Cohort. *Thyroid* **26** 1215-1224. (http://doi.org/10.1089/thy.2016.0016)
268 Macchia PE, Lapi P, Krude H, Pirro MT, Missero C, Chiovato L, Souabni A, Baserga M, Tassi V, Pinchera A, *et al.*
269 1998 PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. *Nat Genet* **19** 83-86.

(http://doi.org/10.1038/ng0598-83)

Mansouri A, Chowdhury K & Gruss P 1998 Follicular cells of the thyroid gland require Pax8 gene function. *Nat Genet* **19** 87-90. (http://doi.org/10.1038/ng0598-87)

Meeus L, Gilbert B, Rydlewski C, Parma J, Roussie AL, Abramowicz M, Vilain C, Christophe D, Costagliola S & Vassart G 2004 Characterization of a novel loss of function mutation of PAX8 in a familial case of congenital hypothyroidism with in-place, normal-sized thyroid. *J Clin Endocrinol Metab* **89** 4285-4291. (http://doi.org/10.1210/jc.2004-0166)

Narumi S, Araki S, Hori N, Muroya K, Yamamoto Y, Asakura Y, Adachi M & Hasegawa T 2012 Functional characterization of four novel PAX8 mutations causing congenital hypothyroidism: new evidence for haploinsufficiency as a disease mechanism. *Eur J Endocrinol* **167** 625-632. (http://doi.org/10.1530/EJE-12-0410)

Narumi S, Muroya K, Asakura Y, Adachi M & Hasegawa T 2010 Transcription factor mutations and congenital hypothyroidism: systematic genetic screening of a population-based cohort of Japanese patients. *J Clin Endocrinol Metab* **95** 1981-1985. (http://doi.org/10.1210/jc.2009-2373)

Narumi S, Yoshida A, Muroya K, Asakura Y, Adachi M, Fukuzawa R, Kameyama K & Hasegawa T 2011 PAX8 mutation disturbing thyroid follicular growth: a case report. *J Clin Endocrinol Metab* **96** E2039-2044. (http://doi.org/10.1210/jc.2011-1114)

Ohno M, Zannini M, Levy O, Carrasco N & di Lauro R 1999 The paired-domain transcription factor Pax8 binds to the upstream enhancer of the rat sodium/iodide symporter gene and participates in both thyroid-specific and cyclic-AMP-dependent transcription. *Mol Cell Biol* **19** 2051-2060.

Plachov D, Chowdhury K, Walther C, Simon D, Guenet JL & Gruss P 1990 Pax8, a murine paired box gene expressed

290 in the developing excretory system and thyroid gland. *Development* **110** 643-651.

291 Ramos HE, Carre A, Chevrier L, Szinnai G, Tron E, Cerqueira TL, Leger J, Cabrol S, Puel O, Queinnec C*, et al.* 2014

292 Extreme phenotypic variability of thyroid dysgenesis in six new cases of congenital hypothyroidism due to PAX8

293 gene loss-of-function mutations. *Eur J Endocrinol* **171** 499-507. (http://doi.org/10.1530/EJE-13-1006)

294 Reva B, Antipin Y & Sander C 2011 Predicting the functional impact of protein mutations: application to cancer

295 genomics. *Nucleic Acids Res* **39** e118. (http://doi.org/10.1093/nar/gkr407)

296 Schimmenti LA 2011 Renal coloboma syndrome. *Eur J Hum Genet* **19** 1207-1212.

297 (http://doi.org/10.1038/ejhg.2011.102)

298 Schwarz JM, Rodelsperger C, Schuelke M & Seelow D 2010 MutationTaster evaluates disease-causing potential of

299 sequence alterations. *Nat Methods* **7** 575-576. (http://doi.org/10.1038/nmeth0810-575)

300 Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN & Gaunt TR 2013 Predicting the

301 functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum*

302 *Mutat* **34** 57-65. (http://doi.org/10.1002/humu.22225)

303 Srichomkwun P, Admoni O, Refetoff S & de Vries L 2016 A Novel Mutation (S54C) of the PAX8 Gene in a Family

304 with Congenital Hypothyroidism and a High Proportion of Affected Individuals. *Horm Res Paediatr* **86** 137-142.

305 (http://doi.org/10.1159/000445891)

306 Thomas PQ, Brown A & Beddington RS 1998 Hex: a homeobox gene revealing peri-implantation asymmetry in the

307 mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125** 85-94.

308 Thusberg J, Olatubosun A & Vihinen M 2011 Performance of mutation pathogenicity prediction methods on missense

309 variants. *Hum Mutat* **32** 358-368. (http://doi.org/10.1002/humu.21445)

Tonacchera M, Banco ME, Montanelli L, Di Cosmo C, Agretti P, De Marco G, Ferrarini E, Ordookhani A, Perri A, Chiovato L*, et al.* 2007 Genetic analysis of the PAX8 gene in children with congenital hypothyroidism and dysgenetic or eutopic thyroid glands: identification of a novel sequence variant. *Clin Endocrinol (Oxf)* **67** 34-40. (http://doi.org/10.1111/j.1365-2265.2007.02831.x)

Vilain C, Rydlewski C, Duprez L, Heinrichs C, Abramowicz M, Malvaux P, Renneboog B, Parma J, Costagliola S & Vassart G 2001 Autosomal dominant transmission of congenital thyroid hypoplasia due to loss-of-function mutation of PAX8. *J Clin Endocrinol Metab* **86** 234-238. (http://doi.org/10.1210/jcem.86.1.7140)

Zannini M, Avantaggiato V, Biffali E, Arnone MI, Sato K, Pischetola M, Taylor BA, Phillips SJ, Simeone A & Di Lauro R 1997 TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation. *EMBO J* **16** 3185-3197. (http://doi.org/10.1093/emboj/16.11.3185)

Zannini M, Francis-Lang H, Plachov D & Di Lauro R 1992 Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol Cell Biol* **12** 4230-4241.

Zou H, Chai J, Liu S, Zang H, Yu X, Tian L, Li H & Han B 2015 A De novo PAX8 mutation in a Chinese child with congenital thyroid dysgenesis. *Int J Clin Exp Pathol* **8** 11434-11439.

Figure legends

Figure 1

Effects of alanine substitutions in paired domain of PAX8. (A) A schematic diagram showing the secondary structure of the paired domain (Gly^{9} to Gln^{137}) of PAX8. Residues that caused profound loss of function (LOF), moderate LOF and minimal LOF were colored in red, orange and yellow, respectively. Residues with comparable activities with wildtype and high activity were colored in gray and blue, respectively. Two β-sheets and six α-helices are showed as boxes. Bars indicate the locations of previously reported missense *PAX8* mutations. (B-E) The three-dimensional structure of the DNA-binding paired domain and its target DNA (colored in silver), based on the crystal structure data of PAX5–DNA complex. An overall view indicating the three subdomains (B): N-subdomain (C), C-subdomain (D), and the linker polypeptide (E).

Figure 2

ROC curves of seven computational algorithms (FATHMM, MutationAssessor, MutationTaster, PolyPhen-2, PROVEAN, SIFT, and VEST-4) are shown. For each of 132 PAX8 alanine variants, scores were obtained with the seven algorithms. With assuming *in vitro* transactivating capacities as a gold standard test, ROC curves were drawn.

345 **Table 1**

346 **Structure-function relationships of the N-subdomain and C-subdomain**

Classification of the effect	N-subdomain			C-subdomain			*P*[b]
	Total (N)	LOF[a] (N)	%LOF (%)	Total (N)	LOF[a] (N)	%LOF (%)	
Loss of hydrogen bond(s) to DNA	12	10	83	4	3	75	0.47
Loss of van der Waals contact(s) to DNA	0	0	0	1	0	0	NA
Loss of hydrogen bond(s) within PAX8	6	3	50	11	6	55	0.38
Loss of van der Waals contact (within PAX8)	19	14	74	18	11	61	0.20
No effect (“free sidechain”)	16	11	69	11	3	27	0.036

347 a Significant LOF *(i.e.,* relative activity <30% of wildtype)

348 b P values were calculated with Fisher exact test.

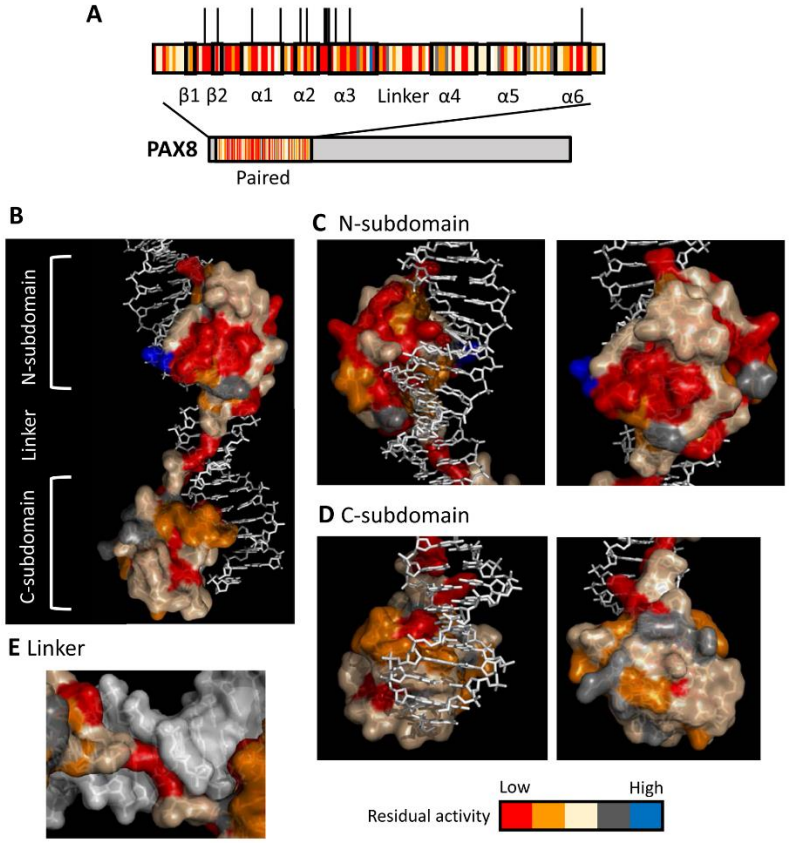
349 Abbreviations: LOF loss of function

350

Table 2

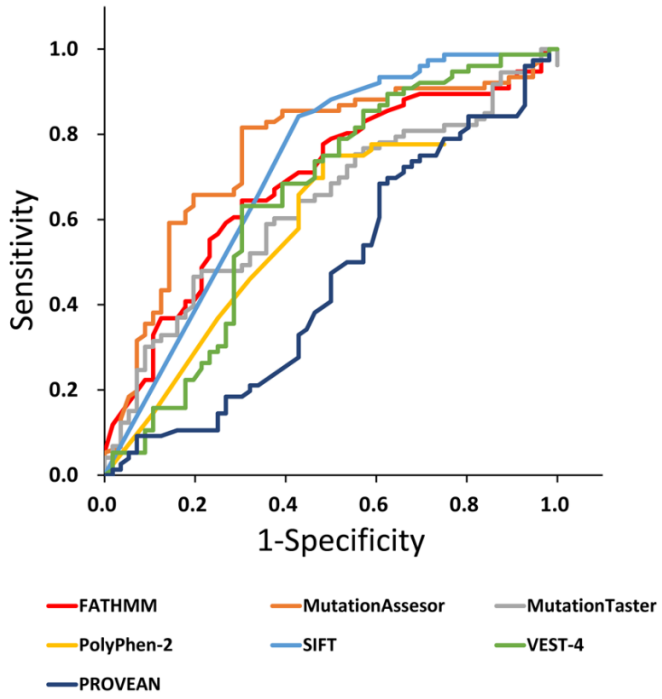
Effects of alanine substitutions on 13 residues that have been affected in human CH patients

Human mutation	Alanine Variant	*TG*-luc activity of Ala variant (%)
Leu16Pro	Leu16Ala	5.8±1.5
Phe20Ser	Phe20Ala	-0.4±2.0
Pro25Arg	Pro25Ala	21.2±2.6
Arg31His	Arg31Ala	-0.9±0.4
Arg31Cys	Arg31Ala	-0.9±0.4
Gln40Pro	Gln40Ala	61.0±1.8
Ile47Thr	Ile47Ala	-1.9±0.9
Ser48Phe	Ser48Ala	22.7±4.2
Arg52Pro	Arg52Ala	0.9±4.9
Ser54Arg	Ser54Ala	-0.6±0.9
Ser54Gly	Ser54Ala	-0.6±0.9
Ser54Cys	Ser54Ala	-0.6±0.9
His55Gln	His55Ala	9.2±1.8
Cys57Tyr	Cys57Ala	16.1±6.7
Leu62Arg	Leu62Arg	-1.4±0.9
Arg133Gln	Arg133Ala	11.9±4.1



355

356 Fig. 1



357

358 Fig. 2